

Article

Not peer-reviewed version

ML-Based Optimal Design of One-Component Ionizable Amphiphilic Janus Dendrimers for Enhanced Dendrimersome Nanoparticle-Mediated mRNA Delivery

[Joshua Kim](#) and [Sungwoo Yang](#) *

Posted Date: 30 October 2025

doi: 10.20944/preprints202510.2331.v1

Keywords: IAJD; DNP; mRNA; delivery; encapsulation; transfection efficiency; luminescence; machine learning; optimization; prediction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

ML-Based Optimal Design of One-Component Ionizable Amphiphilic Janus Dendrimers for Enhanced Dendrimersome Nanoparticle-Mediated mRNA Delivery

Joshua Kim ¹ and Sungwoo Yang ^{2,*}

¹ McCallie School, Chattanooga, TN 37404, USA

² Department of Civil and Chemical Engineering, University of Tennessee at Chattanooga, Chattanooga, TN 37403, USA

* Correspondence: Sungwoo-yang@utc.edu; Tel.: +1-423-425-4366

Abstract

Background/Objectives: Ionizable lipid nanoparticles (LNPs) are the mainstream delivery mechanisms for mRNA vaccines. However, LNPs are limited in their mRNA transfection efficiency (TE) into target cells. Dendrimersome nanoparticle (DNP) delivery systems, developed using ionizable amphiphilic Janus dendrimers (IAJDs), were designed to overcome the limitations of earlier approaches. Researchers have found this alternative promising due to their comparatively simple, repeating one-component structure and enhanced stability. This study sought to clarify the impact of particular IAJD structural components on mRNA TE and develop novel IAJD candidates for maximum predicted TE. **Methods:** Structural constituents (hydrophilic, ionizable amine, & hydrophobic regions) were systematically defined & encoded for computational analysis. Luciferase-induced luminescence was used as a quantitative metric for mRNA transfection. TE prediction models were built using several machine learning algorithms, and the model using eXtreme Gradient Boosting was selected. This prediction model overcame imbalanced datasets and this model was used to find the optimal IAJD designs and formulation conditions. **Results:** The IAJD optimization process ultimately yielded three novel optimized IAJD candidates and one of existing IAJDs, surpassing previously identified IAJDs. **Conclusions:** To our knowledge, this study presents the first large-scale computational investigation of IAJD structural optimization using machine learning. The design of IAJD is the primary factor that influences mRNA TE, but there are other impacting factors and more work is needed. This study highlights the potential of ML-driven IAJD optimization. Combined with high-throughput *in vitro* assays, this method could significantly accelerate mRNA therapeutics development with an improved delivery mechanism.

Keywords: IAJD; DNP; mRNA; delivery; encapsulation; transfection efficiency; luminescence; machine learning; optimization; prediction

1. Introduction

mRNA vaccines have emerged as an effective alternative to traditional vaccine platforms. mRNA vaccines function by encoding a viral protein that triggers an immune response. Once expressed in host cells, the protein is recognized as foreign, prompting the immune system to generate antibodies that protect against future infections. They achieve this by recognizing and binding to specific pathogens, such as viruses. This attachment marks the pathogens for destruction.

mRNA vaccines have numerous advantages, but they also have limitations that need to be overcome. For mRNA to be therapeutically effective, it must reach the target cell. Once delivered, it must produce a sufficient amount of the desired protein. However, mRNA is vulnerable to

degradation by ribonucleases and other factors in the extracellular environment. Therefore, mRNA needs protection from degradation by nucleases and shields its negative charge. Due to the need for mRNA to be structurally protected, its intracellular delivery is the most challenging aspect of mRNA therapeutics [1].

Direct injection of mRNA is one way to deliver it to target tissues; however, in physiological conditions, it is susceptible to degradation or enzymatic digestion by ribonucleases on its own. Lipid-based carriers, also known as lipid nanoparticles (LNPs), are another method for delivering mRNA to target tissues, and they have emerged as the fastest and most efficient tool for combating infectious diseases. LNP was used for COVID-19 mRNA vaccines [1].

LNPs have proven effective at delivering mRNA to specific tissues. LNP-based mRNA vaccines have proven to be a quick and effective vaccination strategy. However, they need to optimize safety and increase efficacy. Four-component LNPs contain ionizable lipids (IL), phospholipids, cholesterol for mechanical properties, and PEG-conjugated lipid for stability, and they represent the current leading nonviral vectors for mRNA. However, the four-component LNPs have major limitations [2,3]. First, the distribution of its four components in the LNP is unknown during the synthesis process. The segregation of the neutral IL as an oil phase in the core of the LNPs is responsible for very low mRNA transfection efficiency (1-2%). Another limitation is the "PEG dilemma," where PEG increases the circulation time in the blood but decreases gene expression. Moreover, LNPs tend to be stable only at very low temperatures.

While LNPs are widely used, the dendrimersome nanoparticle (DNPs) made from ionizable amphiphilic Janus dendrimers (IAJDs) has emerged as a one-component alternative with targeted delivery capabilities, offering advantages over the traditional four-component lipid nanoparticle (LNP) mRNA delivery systems. DNPs offer several advantages over traditional mRNA delivery agents [2,3]. With their self-assembly capabilities, DNPs can be synthesized in fewer steps than multi-component LNPs, potentially enabling more efficient large-scale production. They co-assemble with mRNA through a simple ethanol injection method, avoiding the complex microfluidic technology required for LNPs. Another advantage is that encapsulating mRNA with IAJDs is fast and simple, allowing for rapid screening of new IAJD properties. Moreover, IAJDs can be designed to target specific organs, such as the lungs, liver, spleen, and lymph nodes, by modifying their hydrophilic and hydrophobic components. Additionally, IAJDs exhibit stability at 5°C, a significant advantage over LNPs, which require low-temperature storage at -70°C. This stability simplifies handling and storage, making IAJDs more practical for widespread use. Transfection efficiency (TE) is another advantage of IAJDs. While traditional LNPs have shown low TE (1-2%), IAJDs have the potential to improve upon this, having demonstrated higher transfection efficiency *in vitro* and *in vivo* compared to traditional delivery systems, making them effective for mRNA delivery. This whole mRNA delivery process for LNP or DNP is illustrated in **Figure 1**.

Recent research has explored machine learning (ML) approaches to optimize the design of mRNA delivery systems. ML models have shown promise in predicting LNP TE and mRNA expression levels, potentially accelerating the development of effective delivery agents [4] and optimizing LNP design for mRNA delivery [5,6]. In LNP-mediated mRNA delivery, a significant bottleneck is the identification of new ionizable lipid, which is a crucial component of LNPs [7]. ML combined with combinatorial chemistry offered a powerful approach to accelerate the discovery of novel lipids [8]. For example, the AI-guided Ionizable Lipid Engineering (AGILE) platform integrated deep learning with combinatorial chemistry to streamline the iterative development of ionizable lipids for LNP-mediated mRNA delivery and this platform enabled the rapid design, synthesis, and evaluation of new ionizable lipids, selecting from a library of over 10,000 candidates for mRNA delivery in muscle and immune cells [8]. Studies have employed ML algorithms to reduce the burden of screening ionizable lipids by learning from existing experimental data [9]. This demonstrates the potential of ML to accelerate new ionizable lipid development by simplifying the screening efficiency by computationally prioritizing LNP candidates for experimental validation [9,10].

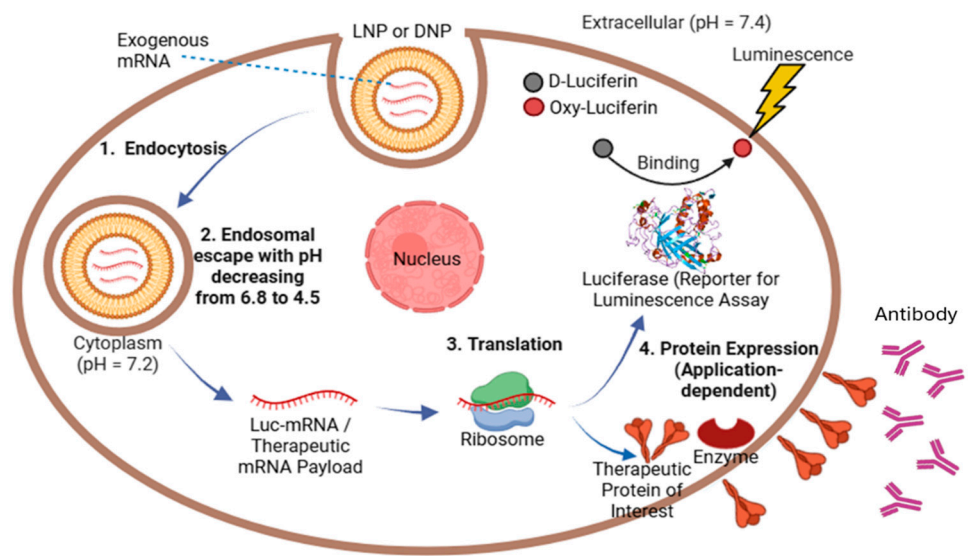


Figure 1. mRNA delivery process for LNP or DNP. Once injected, LNP or DNP approach the corresponding cells and get encapsulated via endocytosis. The endocytosis of LNPs or DNPs deposits them into endosomes and releases the mRNA into the cytoplasm, which allows the ribosome to generate the new proteins. If therapeutic mRNA was released, therapeutic protein is generated and an antibody is formed. If Luc-mRNA was released, luciferase is generated. Luciferase-induced luminescence is widely used as a proxy to measure TE.

ML has been used to predict key performance measures for mRNA delivery systems. ML models can predict translational efficiency (TE) with high accuracy, potentially accelerating the development of effective mRNA delivery systems [4]. One study demonstrated that a multilayer perceptron (MLP) model achieved a classification accuracy of 98% on a test set of unseen LNPs from a curated dataset of 622 LNPs, effectively predicting transfection [10]. A comprehensive ML framework was developed to predict the activity and cell viability of LNPs for nucleic acid delivery, leveraging data from 6,454 LNP formulations reported across 21 independent studies [11]. Furthermore, random forest regression models trained with 314 features have been used to analyze 213 LNPs and predict mRNA expression efficiency after intradermal injection in mice. These models identified phenol as a dominant substructure affecting mRNA encapsulation and expression and highlighted the impact of specific phospholipids, N/P ratio, and mass ratio, as well as carbon chain length, on mRNA delivery efficacy and LNP stability [12].

ML also plays a significant role in optimizing the manufacturing processes of LNPs, which typically involve an upstream process for preparing LNPs and a downstream process for purification [13]. In one study, an ML technique identified ethanol concentration (flow rate ratio), buffer pH, and total flow rate as process parameters that significantly affected the particle size and encapsulation efficiency of LNPs. Based on these findings, this study derived manufacturing conditions for LNPs with different particle sizes (approximately 80 and 200 nm) [13]. This study optimized mRNA-LNP vaccine quality by considering manufacturing and lipid mix ratios [5].

ML is also being applied to generate biological insights into the mechanism of LNP delivery of mRNA, particularly for MC3-LNP [14]. The workflow, termed Advanced Cellular and Endocytic profiling for Intracellular Delivery (ACE-ID), used a cell-based imaging assay and perturbation of 178 targets relevant to intracellular trafficking to identify corresponding effects on functional mRNA delivery. In this study, ML determines key features correlating with enhanced delivery, identifying fluid-phase endocytosis as a productive cellular entry route. It has allowed for reengineering MC3-LNP to target micropinocytosis, significantly improving mRNA delivery *in vitro* and *in vivo* [14].

The application of ML also facilitates the rational design of LNPs for enhanced mRNA delivery, enabling tailored formulations for specific therapeutic applications and cell types [8,12]. In one study, ML-assisted design was used to create immunomodulatory LNPs for delivering mRNA to repolarize hyperactivated microglia, a promising strategy for treating neurodegenerative and autoimmune

disorders [15]. This study employed supervised ML classifiers to investigate 216 LNP formulations with varying lipid compositions, N/P ratios, and hyaluronic acid (HA) modifications, assessing transfection efficiency of eGFP mRNA in BV-2 murine microglia cells under different immunological states [15]. Furthermore, ML has been applied to engineer PoLixNano nanoparticles, which are nebulized mRNA therapeutics designed to overcome delivery barriers [16]. This study developed a quantitative structure-mRNA transfection prediction (QSMTP) model, leveraging physicochemical properties and inhaled mRNA transfection profiles of a chemically diverse library of polymeric components to identify top-performing amphiphilic-copolymers from over 10,000 candidates and suggested that their mucus-penetrating ability outweighs shear force resistance in contributing to efficient mRNA transfection [16].

A variety of ML algorithms and models have been successfully applied in LNP research. Light gradient boosting machine (LightGBM) was used for predicting LNP potency and identifying key structural features influencing transfection efficiency [9]. Extreme gradient boosting (XGB) was applied in optimizing LNP manufacturing processes to identify critical parameters like ethanol concentration, buffer pH, and total flow rate affecting particle size and encapsulation efficiency [5,13]. Multilayer perceptron (MLP) was used to achieve high classification accuracy (98%) in predicting the transfection efficiency of unseen LNPs [10]. It was also the best-performing model (weighted F1-scores ≥ 0.8) in predicting transfection efficiency and phenotypic changes for immunomodulatory LNP design [15]. In one study, random forest (RF) and gradient boosting (GB) models provided the most accurate predictions (over 90% accuracy) for LNP activity and cell viability in large-scale datasets, particularly when combined with molecular descriptors [17]. In another study, Bayesian optimization was used in conjunction with XGB to derive optimal manufacturing conditions for LLNPs with specific particle sizes [5,13]. Also, deep learning was integrated into platforms like AGILE for the comprehensive *in silico* screening of ionizable lipids [8]. These advancements offer promising avenues for improving mRNA delivery in vaccine development. Despite the potential benefits of the ML approach, it hasn't been applied to IAJD research yet. This study utilizes current *in vitro* IAJD data to biochemically optimize an IAJD-based mRNA delivery agent with maximum transfection efficiency (TE) using an ML approach. We used the data to train and build ML models and selected the best-performing prediction model after hyperparameter tuning for each. Using the selected model, we identified the optimal IAJD designs and DNP formulation conditions that provide the best predicted TE. This project is novel since, to our knowledge, this is the first ML application to the IAJD study, not to mention that IAJD itself is a new area of research. This study proposes a framework for future ML applications to IAJD research for the first time. Additionally, it highlights the potential of the ML approach to accelerate the development of mRNA-based therapeutics and overcome the limitations of current mRNA delivery agents.

2. Materials and Methods

This study manually extracts data from recent IAJD-related papers [2,3,18–21]. The extracted data were curated and condensed into 296 records. Computational optimization for IAJD design has not yet been implemented; however, this study aims to address this gap. Variables were defined to describe the molecular structure of all possible IAJDs, allowing ML models to read existing IAJD designs and generate new ones. To define the structural variables, the entire structure of IAJD is broken down into three parts, as shown in **Figure 2**: the hydrophilic part (comprising hydrophilic acids), the ionizable amine (located at the end of each hydrophilic branch), and the hydrophobic part (comprising hydrophobic benzyl amines, hydrophobic acids, or aliphatic alcohols).

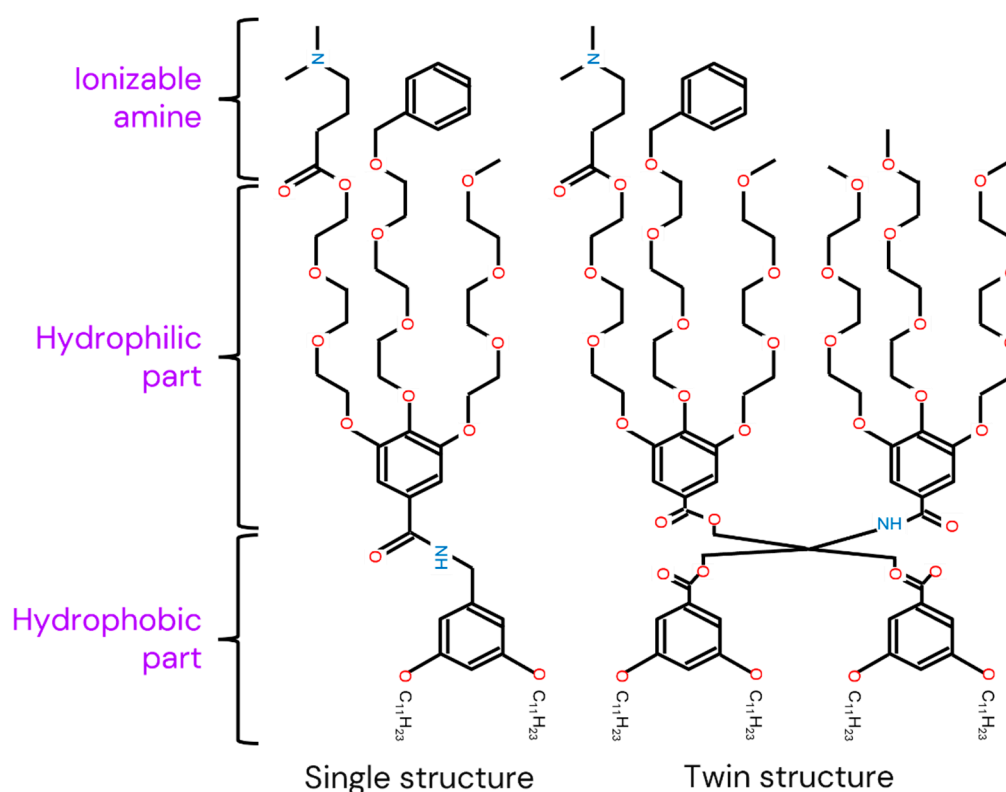


Figure 2. Components of IAJD structure. Every IAJD has three major components: hydrophobic part, hydrophilic part, and ionizable amine. In case of twin structures, at least one of the twins has an ionizable amine.

This study created two datasets: one with IAJD-specific data and another with trial-specific data. IAJD-specific *in vitro* laboratory data from IAJD1 through IAJD311 were manually extracted from IAJD papers published in 2021-2025. Some IAJD numbers between 1 and 311 are missing in publications, though. Variables and their possible values in the IAJD-specific dataset are: IAJD #, pKa (not used in this project), $Ve_{\mu L}$ (not used in this project), SMILES string for the molecular structure (added separately. Used for 3D model generation and chemical properties calculation.), 27 defined structural variables (added separately for prediction model building).

Variables and their possible values in the trial-specific dataset are: IAJD #, Buffer ([1,2] 1 for Acetate, 2 for Citrate), pH_of_buffer ([3.0, 4.0, 4.4, 4.8, 5.2, 5.6] (10 mM)), CIAJD ([0.25, 0.50, 0.75, 1.00, 2.00, 3.00, 4.00, 5.00, 6.00, 7.00, 8.00] (mg/mL) concentration of IAJD when formulating DNP), Cluc_mRNA ([0.05, 0.10] (mg/mL) concentration of luc mRNA when co-assembling with IAJD), Luminescence (the mean of measured luminescence).

These two data sets were combined using IAJD # as the key in Microsoft Access and combined dataset had 413 records. The records without luminescence value were excluded and the data set was cleaned using Microsoft Excel, as it's easier due to the relatively small dataset size. A total of 296 records were used for building the ML model.

Not every feature (column) in the dataset was used for building the ML model. For example, each IAJD's chemical structure was converted into a machine-readable format, i.e., SMILES, and included in the data set but was not used for ML model building in this study. They are saved for molecular characterization and future research projects. For the prediction model's input, 27 variables describe particular chemical components of the IAJD molecular structure, while an additional four variables describe DNP formulation conditions. Details about the structural variables and formulation condition variables are explained in Appendix B. *In vitro* luminescence, which serves as a measure of mRNA transfection efficiency (TE), functions as the dependent variable. Independent & dependent variables are summarized in **Table 1**.

Table 1. Variables in the dataset and the reason for selection.

Variable Type	Variable Name	Reason for Selection
Independent	Buffer Type	Critical factor in DNP formulation (non-structural)
	pH of Buffer	
	Concentration of IAJD	
	Concentration of luciferase mRNA	Classification of overall structure
	Single or Twin Structure	
	Type / Size of Hydrophilic Part	
	Type of Ionizable Amine	
	Type / Size of Hydrophobic Part	Classification of a major IAJD component
	Linker Type	
Dependent	Luminescence (AU)	Measure of mRNA TE

Using the dataset, several machine learning algorithms known for their predictive performance and robustness are selected and trained. For the prediction model from each ML algorithm, the hyperparameters are tuned, and each prediction model is optimized. Multilayer Perceptron (MLP), random forest (RF), gradient boosting (GB), histogram gradient boosting (HGB), and extreme gradient boosting (XGB) were selected in this project. The optimized prediction models are compared, and the best-performing model is selected based on R2 and RMSE values. Then, prediction models undergo cross-validation to assess their consistency in performance and confirm that the best-performing model consistently outperforms the other models. Once the best-performing prediction model is finalized, it is used to determine the optimal IAJD designs and formulation conditions that yield the highest predicted luminescence value. Once the optimal IAJD designs are found, their synthetic accessibility and other characteristics are analyzed. The whole modeling process is summarized in **Figure 3**.

The optimization of the IAJD design and formulation conditions is a separate process from building the best-performing machine learning (ML) model. Once the best prediction model is built, it is used to search for the optimal IAJD design and formulation conditions within the predefined solution space. First, the process generates a new IAJD design and a DNP formulation condition from the pre-defined solution space. The solution space is pre-defined based on the possible values of each variable in the dataset, which means that the search is limited to that pre-defined solution space. Then, the previously selected prediction model predicts the mRNA TE for the candidate IAJD design and the corresponding formulation condition. The newly predicted mRNA TE is compared with the current optimal. If the newly predicted mRNA TE is higher than the current optimal, then the optimal is updated. Otherwise, the optimal is not updated. Then, the process generates another new candidate IAJD design and formulation condition, and repeats the process until all candidates in the solution space have been tested.

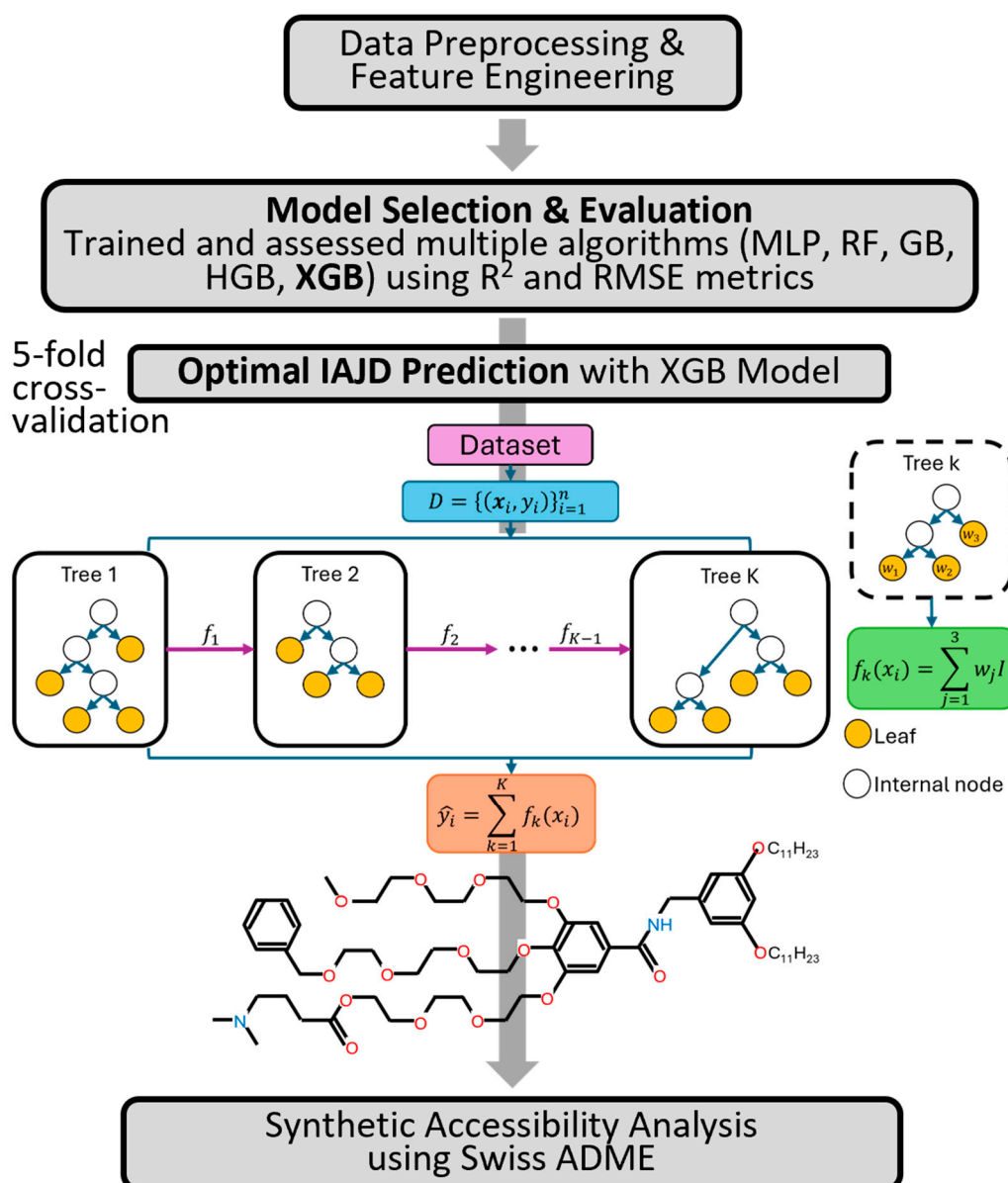


Figure 3. Modeling process. After the data is preprocessed, it is used to train multiple ML algorithms. Prediction models are optimized through hyperparameter-tuning and compared. The best-performing model is selected and undergoes cross-validation to assess its robustness. Then, the model is used to find the optimal IAJD design and formulation condition.

3. Results

After the five different prediction models based on the five ML algorithms were trained using the dataset. The hyperparameters of each prediction model were tuned to optimize each model. The resulting hyperparameter values for each prediction model are summarized in Appendix A. Multi-layer Perceptron (MLP) regressor, Random Forest (RF) regressor, Gradient Boosting (GB) regressor, Histogram Gradient Boosting (HGB) regressor, and Extreme Gradient Boosting (XGB) regressor were evaluated and the top four models were compared as shown in graphs in **Figure 4**. Scatter diagrams in **Figure 4** compare the predicted luminescence values and the observed mean luminescence values for each prediction model. The red dotted lines represent ideal lines, and the data points are mapped onto these lines when the predicted luminescence value matches the observed mean luminescence value perfectly. The graphs show that the luminescence data is somewhat polarized. The data points are densely populated in the 0 to 1 million range, and they are less densely populated in the range above 4 million. They show almost no data points between 1 million and 4 million. Among the tested

prediction models, XGB outperforms the others based on the R^2 value. XGB noticeably outperforms others in both the low and high ranges. XGB is known for its scalable and highly accurate implementation of GB, which pushes the limits of computing power for boosted tree algorithms. XGB is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable [22]. It is also known for its capability to overcome imbalanced datasets. Combined with its L1 and L2 regularization and its ability to effectively analyze multiple variables, it is able to perform accurately [23].

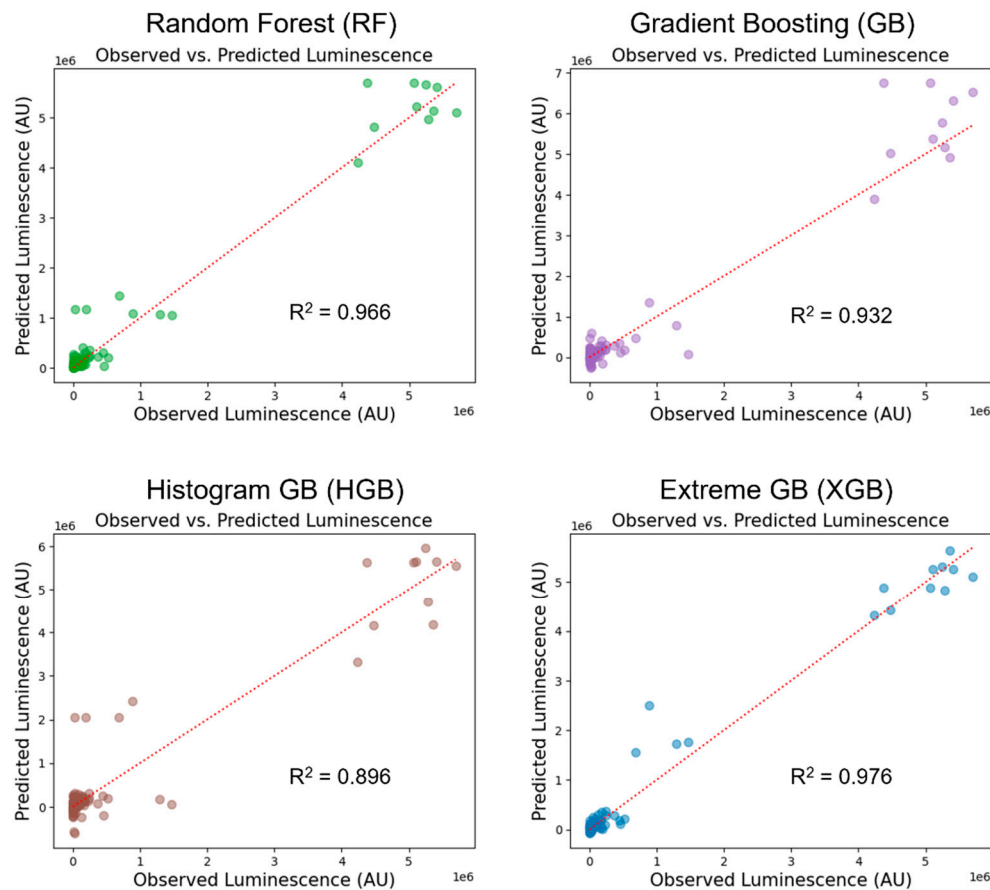


Figure 4. Comparison of ML models tested.

Prediction models were also compared based on Root Mean Squared Error (RMSE) and R-squared (R^2) as shown in **Figure 5**. Lower RMSE indicates a smaller prediction error, and higher R^2 indicates a stronger correlation between the prediction and the actual value. XGB outperforms other models in both RMSE and R^2 . XGB shows the lowest average prediction error (i.e., RMSE) and the highest prediction capability (i.e., R^2).

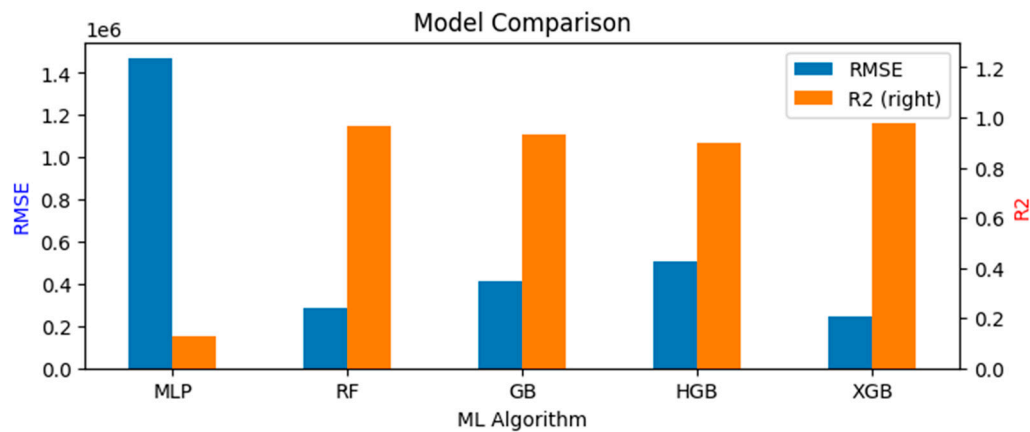


Figure 5. Model comparison by RMSE and R². XGB shows the lowest prediction error (RMSE) and the highest prediction accuracy (R²).

The results indicated that XGB outperforms other models, but cross-validation (5-fold) was also conducted to confirm that XGB consistently outperforms other models. It demonstrates the consistency of their prediction performance when each of the five folds of the dataset is used in turn for validation. Cross-validation was conducted for all the tested models, as shown in **Figure 6**. Error bars in the figure show the range of $\pm$ standard deviation of measured luminescence. XGB outperforms the other ML models in terms of the lowest RMSE and standard deviation by a relatively slim margin. The cross-validation result does not indicate that XGB is significantly better than other models, but it does suggest that XGB is at least not inferior to the others.

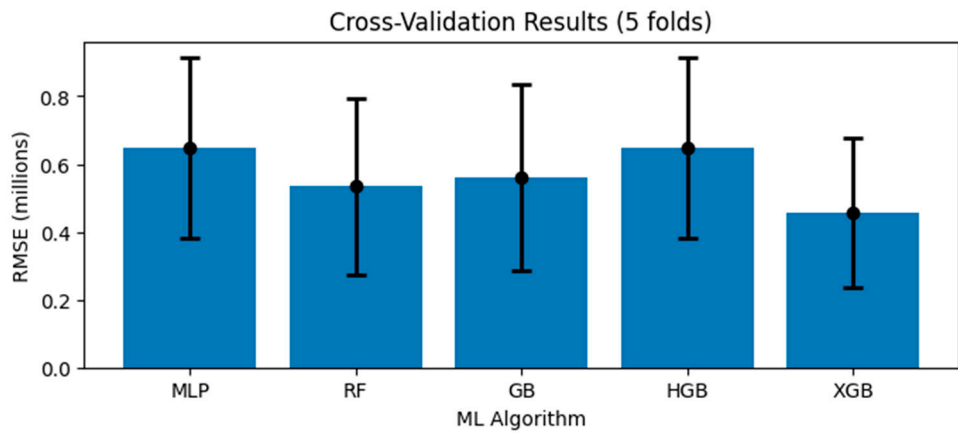


Figure 6. Cross-validation of models (5 folds). On average, XGB exhibits the lowest prediction error, and it also has the lowest variation in prediction error.

Using the XGB-based prediction model, this study searched the pre-defined solution space for the optimal IAJD design and formulation condition. The solution space has been defined as the combination of structural variables and DNP formulation conditions based on the observed values in the dataset. Considering the possible values of each of 31 variables, the initial size of the design space was 1.015×10^{23} . In this solution space, one round of the comprehensive solution search will take 3.6×10^{12} years, i.e., almost 4 trillion years, assuming 900 prediction model calculations per second. Due to the astronomically large size of the solution space, unlikely values of each variable were excluded from the solution space, and invalid combinations were excluded through the search algorithm. All the twin structures in the dataset underperformed compared to single structures, and all the twin structures were excluded from the solution space. With the reduced design space and

improved search algorithm, one round of the solution space search took a more reasonable amount of time: approximately 6 hours.

After the optimal solution search process, four optimal IAJD designs and the optimal formulation condition were identified by the search algorithm using the XGB-based prediction model, as shown in **Figure 7**. The optimal DNP formulation condition is the same regardless of the optimal IAJD design. All four optimal solutions have the same predicted luminescence value of 7,108,495.0 (AU). This is almost as good as the best performing existing IAJD, since 7,317,328 is the highest luminescence value in the training set. The only difference among the four optimals is the HA1_3 variable value, which corresponds to the chemical group located on the third hydrophilic chain. Appendix D presents the solution space and the optimal solutions identified by the XGB-based prediction model.

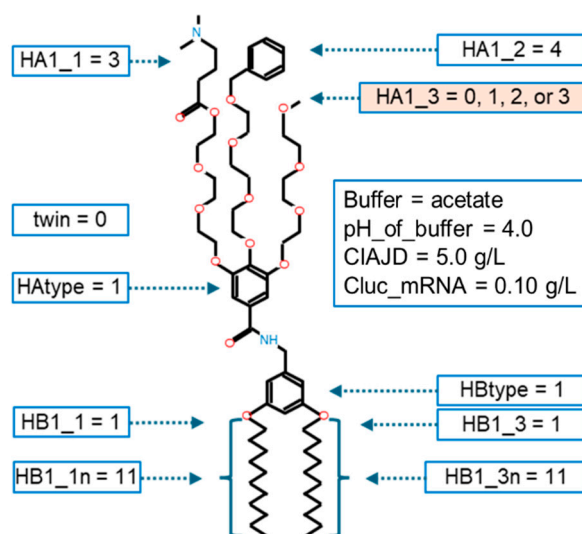


Figure 7. Optimal IAJD designs. Four optimal IAJD designs and formulation conditions have been found. All the optimal solutions share the same values for every variable except for HA1_3.

The variables in **Figure 7** are defined as follows: HAtype (Type of hydrophilic domain), HA1_X (Type of ionizable amine at Xth branch of a type 1 hydrophilic domain), HBtype (Type of hydrophobic domain), HB1_Y (Type of hydrophobic tail at Yth point of type 1 hydrophobic domain), and HB1_Yn (Length of tail at Yth point of type 1 hydrophobic domain).

This study found that the four IAJD designs have different types of ionizable amine at the tip of the 3rd branch of the type 1 hydrophilic domain. Appendix C explains more details about the four optimal solutions. The four optimal IAJD designs are presented in **Figure 8**. The 4th optimal IAJD design is the same as IAJD 22, which is one of the best-performing existing IAJD designs. It is notable that all the optimal IAJD designs are structurally similar to one of the current best-performing IAJDs. IAJD 22 has the highest observed in vitro luminescence value among existing IAJDs, and the prediction model indicates that the candidates close to the top existing IAJD will perform best.

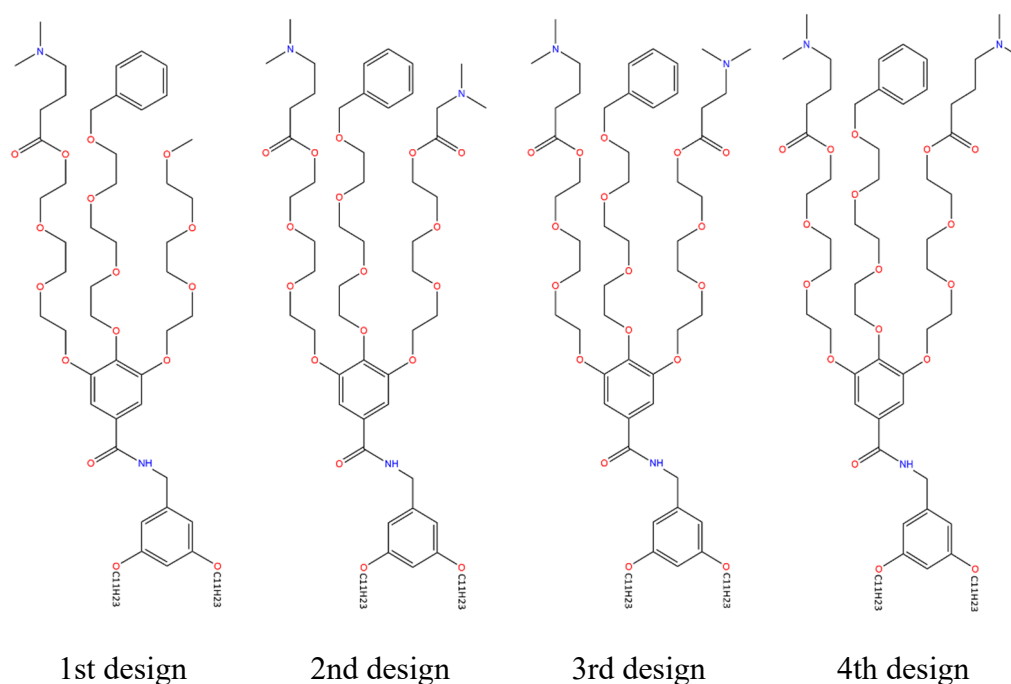


Figure 8. Molecular structures of the optimal IAJD designs. Only the ionizable amine at the tip of the third branch of the hydrophilic part is different among the four optimal designs.

4. Discussion

This study shows promising results, but it has limitations as well. IAJDs were introduced only four years ago, so there is still limited data on these molecules. In this study, luciferase-induced luminescence is used as a measure of TE because it is widely used as a proxy. However, since it is a proxy and does not directly translate to clinical relevance, the interpretation of the results can be limited, and further work is needed to make the results more beneficial. Twin structures have been excluded from the optimal design search due to the poor performance of existing twin structure IAJDs, but this could potentially eliminate promising candidate designs.

Despite the limitations, the study successfully built a robust machine learning model. The model predicts *in vitro* mRNA luminescence and helps identify the optimal IAJD design and formulation conditions. Machine learning (ML) has not yet been applied to IAJD. In this research project, we first defined the structural variables of IAJD. These variables were then entered into a dataset to train ML models. The trained models were used to generate new candidate designs for optimization purposes. Each of the selected ML algorithms was hyperparameter-tuned for the dataset before comparison. Among the five machine learning models tested, the XGB-based prediction model showed the best performance. Therefore, it was selected as the final prediction model. This selected model was then used to identify the optimal IAJD designs and DNP formulation conditions that maximize the predicted TE.

The results demonstrate the ML model's ability to predict TE (as measured by luminescence), even with a relatively small dataset, and how it can be utilized to identify the optimal IAJD design within the prohibitively large design space. The structural variables of IAJD, defined in this project, set the foundation for further machine learning applications in future IAJD research. There can be many factors affecting the mRNA TE other than the molecular structure of IAJD, but there is no doubt that the molecular structure is one of the most important factors. This project has paved the way for further research in this emerging area.

The results of this project have the potential to help develop and produce safe vaccines quickly, particularly with improved storage capabilities and enhanced accessibility of IAJD-based vaccines, especially in underresourced areas, as illustrated in **Figure 9**. Further study will provide more solid benefits to the broader population as more IAJD data becomes available. The next step is to verify the

result of this project by creating and testing the new IAJD designs under optimal DNP formulation conditions. Since the new IAJD designs are not commercially available, one of the new IAJD designs will be custom-synthesized. The IAJD will be co-assembled with luc-mRNA via ethanol injection to formulate novel DNPs. After verification of the formulated DNPs, the novel DNPs will be evaluated against industry-standard ionizable LNPs in cultured HEK293T cells, with transfection assessed via cell plate reader luminescence measurements.

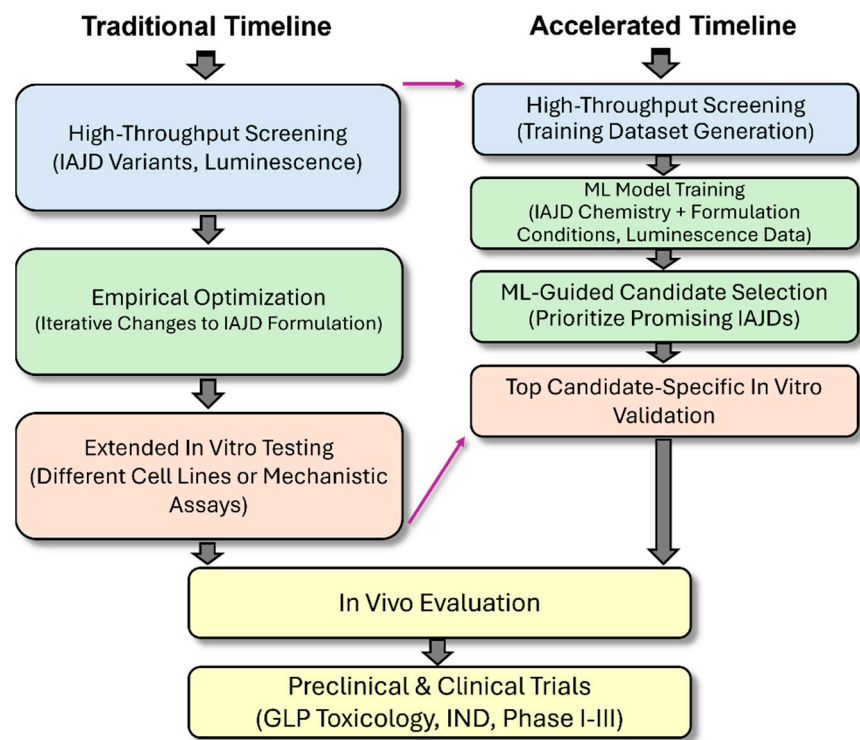


Figure 9. Expected benefit. With the help of the ML model, the timeline for therapeutics development can be shortened, and more lives can be saved at a lower cost and with better accessibility.

Author Contributions: Conceptualization, J.K.; methodology, J.K.; software, J.K.; validation, J.K.; formal analysis, J.K.; investigation, J.K.; resources, S.Y.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, J.K. and S.Y.; visualization, J.K.; supervision, S.Y.; project administration, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

- LNP lipid nanoparticle
- IAJD ionizable amphiphilic Janus dendrimer
- DNP dendrimersome nanoparticle
- mRNA messenger ribonucleic acid
- TE transfection efficiency
- ML machine learning
- MLP multilayer perceptron

RF	random forest
GB	gradient boosting
HGB	histogram gradient boosting
XGB	extreme gradient boosting

Appendix A

Hyperparameter values of prediction models

Through hyperparameter tuning, each prediction model was optimized.

Hyperparameter values of the optimized MLP regressor model:

- alpha=0.019622237860359305
- beta_1=0.9426622228918426
- beta_2=0.981530345935695
- hidden_layer_sizes=125
- learning_rate_init=0.09606142561070556, max_fun=16835
- max_iter=283, momentum=0.7203909798571576
- n_iter_no_change=8, power_t=0.1997727727616871
- tol=0.000609016759626102
- validation_fraction=0.09996090982612142

Hyperparameter values of the optimized random forest (RF) regressor model:

- ccp_alpha=0.08553361303023148
- max_features=0.7751438544438484
- min_impurity_decrease=0.017023303243833054
- min_samples_leaf=1
- min_samples_split=3
- min_weight_fraction_leaf=0.003221536045037288
- n_estimators=90

Hyperparameter values of the optimized gradient boosting (GB) regressor model:

- alpha=0.8680638990931918
- ccp_alpha=0.015750676206814862
- learning_rate=0.1286760958685088
- max_depth=9
- min_impurity_decrease=0.016915940022235176
- min_samples_leaf=1
- min_samples_split=2
- min_weight_fraction_leaf=0.01042397382655489
- n_estimators=76
- subsample=0.5210194554669745
- tol=0.00024704561523662574
- validation_fraction=0.16996781330352345

Hyperparameter values of the optimized histogram gradient boosting (HGB) regressor model:

- l2_regularization=0.029404467922097156
- learning_rate=0.19281183995796564
- max_bins=212
- max_features=0.500763060875326
- max_iter=112
- max_leaf_nodes=13
- min_samples_leaf=10
- n_iter_no_change=12
- validation_fraction=0.26853568630405517

Hyperparameter values of the optimized extreme gradient boosting (XGB) regressor model (manually optimized):

- objective='reg:squarederror'
- max_depth=5
- n_estimators = 802
- learning_rate=0.0105
- reg_lambda=0.17
- reg_alpha=1
- min_child_weight=1
- subsample=1.00
- colsample_bytree=1

Appendix B

Structural variables in the dataset

Variables used in this project to describe the molecular structure of each candidate IAJD are explained as follows:

- twin: [0, 1]
 - 0: single structure
 - 1: twin structure
 - If twin = 0, then Linker = tH* = 0.
- HAtype: [1, 2, 3]. This variable determines the type of hydrophilic part. There are three different types as shown in **Figure B1**.
 - 1: three branches from a Benzene ring
 - 2: a chain of carbons and oxygens with a length n
 - 3: no hydrophilic acid, i.e., ionizable amine only, as illustrated below.
 - If HAtype = 1 or 3, then HA2n = 0.

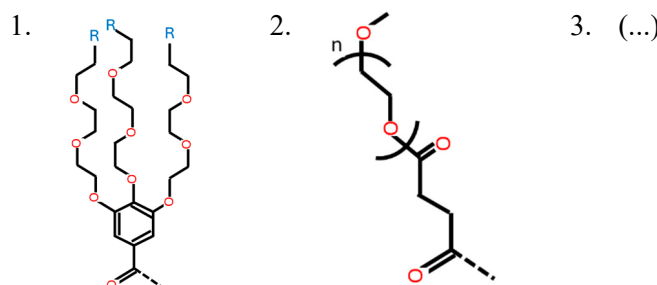


Figure B1. Options for variable HAtype.

- HA2n: [0, 3, 4, 8, 45] This variable determines the length of the carbon and oxygen chain (HAtype 2). This variable is valid only if HAtype = 2. Otherwise, the value of this variable is 0.
- HA1_1: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10] This variable determines the type of ionizable amine at the tip of the 1st branch from the benzene ring (HAtype 1) as shown in Figure B2. There are 11 different types as illustrated in **Figure B3**.
 - If HAtype is not 1, this variable is invalid and its value is 0.
 - If HAtype is 1 and HA1_1 is 0, it means that this branch does not have ionizable amine.

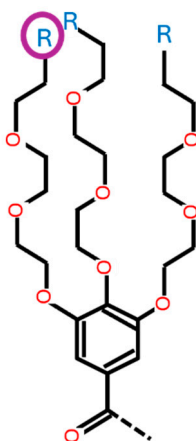


Figure B2. The location associated with HA1_1.

- HA1_2: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
 - This variable determines the type of ionizable amine at the tip of the 2nd branch from the benzene ring (HAtype 1) as shown in **Figure B4**. There are 11 different types as illustrated in **Figure B3**.

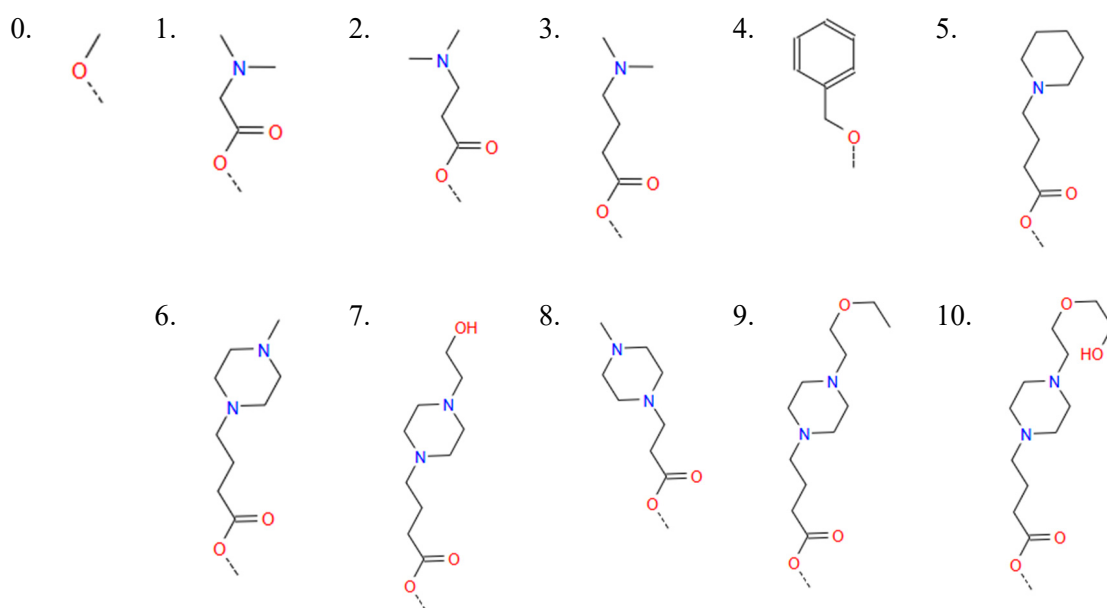


Figure B3. Options for variables HA1_1, HA1_2, and HA1_3.

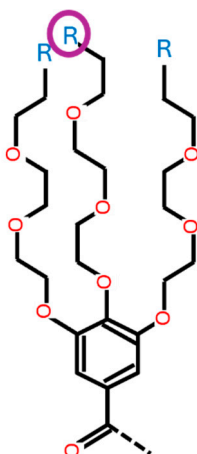


Figure B4. The location associated with HA1_2.

- HA1_3: [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
 - This variable determines the type of ionizable amine at the tip of the 3rd branch from the benzene ring (HAtype 1) as shown in **Figure B5**. There are 11 different types as illustrated in **Figure B3**.

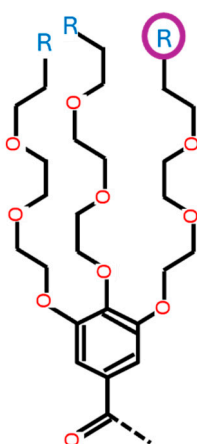


Figure B5. The location associated with HA1_3.

- HBtype: [1, 2, 3, 4] The hydrophilic part of IAJD is connected to a hydrophobic part and this variable determines the type of hydrophobic part as shown in **Figure B6**.
 - 1: Hydrophobic Benzyl Amines (Benzene ring connected with HN)
 - 2: Hydrophobic Acids (Benzene ring connected with oxygen (O) with double links)
 - 3: Benzene ring connected with O
 - 4: Aliphatic Alcohol (three prongs connected with C_nH_{2n+1} ($C_nH_{(2n+1)}$) for each with no Benzene ring) as illustrated below.
 - The dashed line indicates the connection to the hydrophilic part.
 - If HBtype is 1, 2, or 3, $HB4n = 0$.

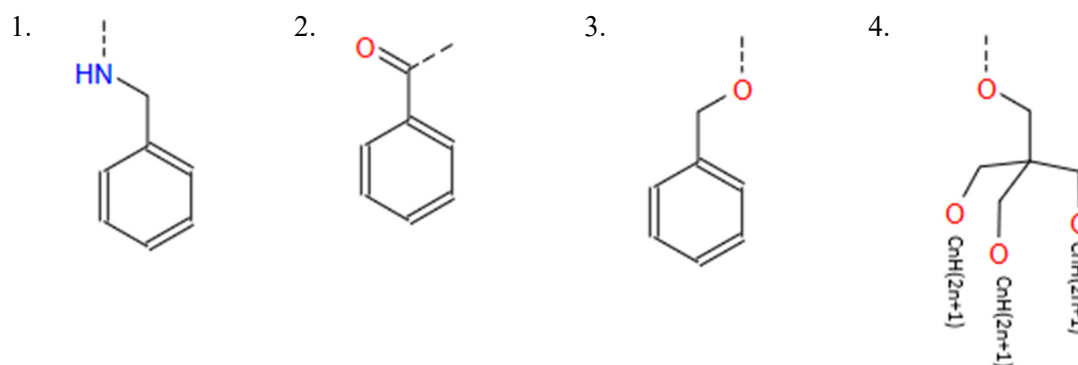


Figure B6. Options for the variable HBtype.

- HB4n: [0, 1, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] This variable determines the length of tail (# of carbons) when HBtype is 4 (Aliphatic Alcohol).
 - This variable is valid only when HBtype is 4.
 - If HBtype is not 4, the value of this variable is 0.
 - If HBtype is 4 and each tail is a two-prong tail as illustrated in Figure B7, the value of this variable is 1.
 - Other values for this variable indicate the tail length (# of carbons) for C_nH_{2n+1} .

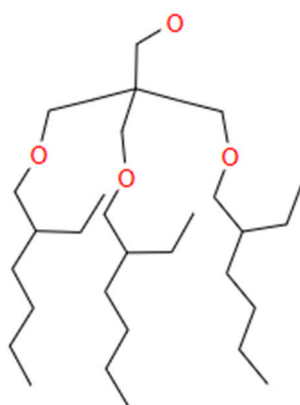


Figure B7. Hydrophobic part when HBtype is 4 and HB4n is 1.

- HB1_1: [0, 1, 2, 3, 4, 5] This variable determines the type of tail at the 1st tail point from the benzene ring of HBtype 1 as shown in **Figure B8**.
 - There are 6 different types as shown in Figure B9.
 - 0 means that there is no tail at the 1st tail point or not applicable.

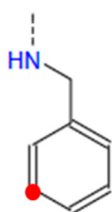


Figure B8. The location associated with HB1_1.

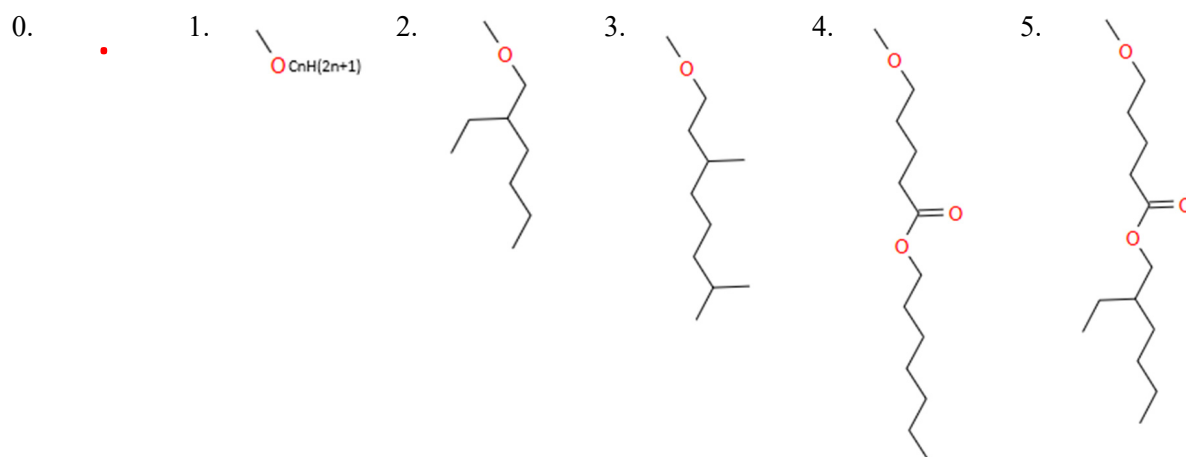


Figure B9. Options for variables HB1_1, HB1_2, and HB1_3.

- HB1_1n: [0, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] This variable indicates the length of the tail at tail point 1 from the benzene ring (BHtype 1), i.e. n value for $C_nH_{(2n+1)}$. 0 means no tail at tail point 1 or N/A.
- HB1_2: [0, 1, 2, 3, 4, 5] This variable determines the type of tail at the 2nd tail point from the benzene ring of HBtype 1 as shown in **Figure B10**.

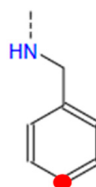


Figure B10. The location associated with HB1_2.

- There are 6 different types as shown in Figure B9.
- 0 means that there is no tail at the 2nd tail point or not applicable.
- HB1_2n: [0, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] This variable indicates the length of the tail at tail point 2 from the benzene ring (BHtype 1), i.e. n value for $C_nH_{(2n+1)}$.
 - 0 means no tail at tail point 2 or N/A.
- HB1_3: [0, 1, 2, 3, 4, 5] This variable determines the type of tail at the 3rd tail point from the benzene ring of HBtype 1 as shown in **Figure B11**.
 - There are 6 different types as shown in Figure B9.
 - 0 means that there is no tail at the 3rd tail point or not applicable.

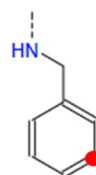


Figure B11. The location associated with HB1_3.

- HB1_3n: [0, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] This variable indicates the length of the tail at tail point 3 from the benzene ring (BHtype 1), i.e., n value for $C_nH_{(2n+1)}$.
 - 0 means no tail at tail point 3 or N/A.

- Linker: [0, 1, 2] This variable indicates the type of linker when IAJD has a twin structure. There are two different types as shown in **Figure B12**.
 - 0: N/A, i.e., not twin.
 - 1: symmetric linker type
 - 2: asymmetric linker type, which has NH on one side.

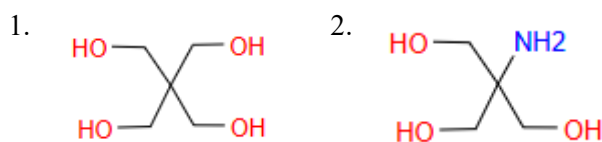


Figure B12. Options for the variable Linker.

- tHAtype: [0, 1, 2, 3] This variable is HAtype for twin. Twin refers to the one with NH as shown in **Figure B13**.
 - 0: This IAJD does not have a twin structure (twin = 0) or N/A. If Linker is asymmetric (Linker = 2), NH side is the twin.

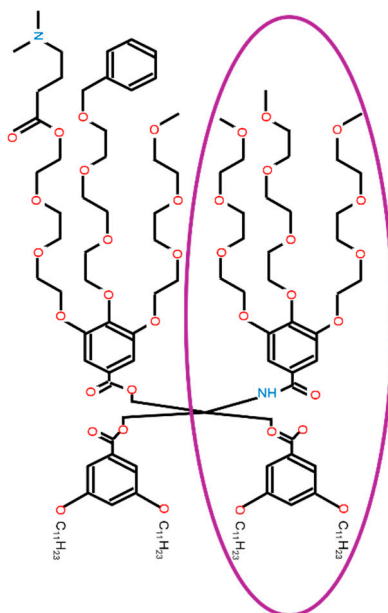


Figure B13. In a twin structure, twin refers to the one with NH.

- tHA2n: [0, 3, 4, 8, 45] This variable is HA2n for twin. 0: This IAJD does not have a twin structure (twin = 0) or N/A.
- tHA1_1: [0, 1, 2, 3, 4, 5] This variable is HA1_1 for twin.
 - 0: This IAJD does not have a twin structure (twin = 0) or N/A.
- tHA1_2: [0, 1, 2, 3, 4, 5] This variable is HA1_2 for twin.
 - 0: This IAJD does not have a twin structure (twin = 0) or N/A.
- tHA1_3: [0, 1, 2, 3, 4, 5] This variable is HA1_3 for twin.
 - 0: This IAJD does not have a twin structure (twin = 0) or N/A.
- tHBtype: [0, 1, 2] This variable is HBtype for twin.
 - Based on the current data set, type 3 or 4 doesn't exist for twin yet.
 - 0: Not have a twin structure (twin = 0) or N/A.
- tHB1_1: [0, 1, 2, 3] This variable is HB1_1 for twin.
 - 0: No tail at the 1st tail point, does not have a twin structure (twin = 0), or N/A.
- tHB1_1n: [0, 8, 9, 10, 11, 12] This variable is HB1_1n for twin.
 - 0: Not have a twin structure (twin = 0) or N/A.

- tHB1_2: [0, 1, 2, 3] This variable is HB1_2 for twin.
 - 0: No tail at the 2nd tail point, does not have a twin structure (twin = 0), or N/A.
- tHB1_2n: [0, 8, 9, 10, 11, 12] This variable is HB1_2n for twin.
 - 0: Not have a twin structure (twin = 0) or N/A.
- tHB1_3: [0, 1, 2, 3] This variable is HB1_3 for twin.
 - 0: No tail at the 3rd tail point, does not have a twin structure (twin = 0), or N/A.
- tHB1_3n: [0, 8, 9, 10, 11, 12] This variable is HB1_3n for twin.
 - 0: Not have a twin structure (twin = 0) or N/A.
- If HAtype = 1, then HA1_1 + HA1_2 + HA1_3 $\geq$ 1.
 - (at least 1 branch from hydrophilic benzene ring)
- If HBtype = 1, then HB1_1 + HB1_2 + HB1_3 $\geq$ 1.
 - (at least 1 tail from hydrophobic benzene ring)

Appendix C

Optimal IAJD designs found

Based on the predicted TE (luminescence intensity), the four best IAJD designs were found from my XGB model:

- TE: 7108495.0, params: (1, 4.0, 5.0, 0.1, 0, 1, 0, 3, 4, 0, 1, 0, 1, 11, 0, 0, 1, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
- TE: 7108495.0, params: (1, 4.0, 5.0, 0.1, 0, 1, 0, 3, 4, 1, 1, 0, 1, 11, 0, 0, 1, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
- TE: 7108495.0, params: (1, 4.0, 5.0, 0.1, 0, 1, 0, 3, 4, 2, 1, 0, 1, 11, 0, 0, 1, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
- TE: 7108495.0, params: (1, 4.0, 5.0, 0.1, 0, 1, 0, 3, 4, 3, 1, 0, 1, 11, 0, 0, 1, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

The result indicates that the four best IAJD designs have the highest predicted luminescence value under a certain DNP formulation condition.

Based on this result, the IAJD design's molecular structure was drawn and converted to SMILES string using OSRA online tool. Synthetic accessibility for each design was generated using SwissADME.

- 1st IAJD design
 - CCCCCCCCCCCCOc3cc(CNC(=O)c2cc(OCCOCCOCCOC)c(OCCOCCOCCOCc1cccc1)c(OCCOCCOCCOC(=O)CCCN(C)C)c2)cc(OCCCCCCCCCCC)c3
 - Synthetic accessibility: 9.46
- 2nd IAJD design
 - CCCCCCCCCCCCOc3cc(CNC(=O)c2cc(OCCOCCOCCOC(=O)CCCN(C)C)c(OCCOCCOCCOCc1cccc1)c(OCCOCCOCCOC(=O)CCCN(C)C)c2)cc(OCCCCCCCCCCC)c3
 - Synthetic accessibility: 10.00
- 3rd IAJD design
 - CCCCCCCCCCCCOc3cc(CNC(=O)c2cc(OCCOCCOCCOC(=O)CCCN(C)C)c(OCCOCCOCCOCc1cccc1)c(OCCOCCOCCOC(=O)CCN(C)C)c2)cc(OCCCCCCCCCCC)c3
 - Synthetic accessibility: 10.00
- 4th IAJD design (= IAJD22)
 - CCCCCCCCCCCCOc3cc(CNC(=O)c2cc(OCCOCCOCCOC(=O)CCCN(C)C)c(OCCOCCOCCOCc1cccc1)c(OCCOCCOCCOC(=O)CCCN(C)C)c2)cc(OCCCCCCCCCCC)c3
 - Synthetic accessibility: 10.00

Four commercially available IAJDs were selected for comparison, and their synthetic accessibility scores were found using SwissADME as follows:

- IAJD 34: synthetic accessibility 10.00
- IAJD 93: synthetic accessibility 6.76
- IAJD 249: synthetic accessibility 5.85

- IAJD 288: synthetic accessibility 7.28
This result indicates that my new IAJDs are harder to synthesize than most of the commercially available IAJDs.

Appendix D

Solution Space and Optimal IAJD from XGB

Table D1. Optimal IAJD design and formulation condition identified by the XGB model.

Variable	Possible values											
Buffer	1	2										
pH_of_buffer	0.25	4.0	4.4	4.8	5.2	5.6						
CIAJD	0.25	0.50	0.75	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	
Cluc_mRNA	0.05	0.10										
twin	0	1										
HAtype	1	2	3									
HA2n	0	3	4	8	45							
HA1_1	0	1	2	3	4	5	6	7	8	9	10	
HA1_2	0	1	2	3	4	5	6	7	8	9	10	
HA1_3	0	1	2	3	4	5	6	7	8	9	10	
HBtype	1	2	3	4								
HB4n	0	1	6	7	8	9	10	11	12	13	14	15
HB1_1	0	1	2	3	4	5						
HB1_1n	0	8	9	10	11	12	13	14	15	16	17	18
HB1_2	0	1	2	3	4	5						
HB1_2n	0	8	9	10	11	12	13	14	15	16	17	18
HB1_3	0	1	2	3	4	5						
HB1_3n	0	8	9	10	11	12	13	14	15	16	17	18
Linker	0	1	2									
tHAtype	0	1	2	3								
tHA2n	0	3	4	8	45							
tHA1_1	0	1	2	3	4	5						
tHA1_2	0	1	2	3	4	5						
tHA1_3	0	1	2	3	4	5						
tHBtype	0	1	2									
tHB1_1	0	1	2	3								
tHB1_1n	0	8	9	10	11	12						
tHB1_2	0	1	2	3								
tHB1_2n	0	8	9	10	11	12						
tHB1_3	0	1	2	3								
tHB1_3n	0	8	9	10	11	12						

- : Values excluded from model building
- : Values included in model building but not selected as best
- : Values common among different optimal solutions
- : 1st design
- : 2nd design
- : 3rd design
- : 4th design (= IAJD22)
- Predicted luminescence: 7,108,495.0 (AU)

References

1. Nitika, J. Wei, and A.-M. Hui, *The Delivery of mRNA Vaccines for Therapeutics*. Life, 2022. **12**(8).
2. Zhang, D., Atochina-Vasserman, E. N., Maurya, D. S., Huang, N., Xiao, Q., Ona, N., Liu, M., Shahnawaz, H., Ni, H., Kim, K., Billingsley, M. M., Pochan, D. J., Mitchell, M. J., Weissman, D., and Percec, V. , *One-Component Multifunctional Sequence-Defined Ionizable Amphiphilic Janus Dendrimer Delivery Systems for mRNA*. Journal of the American Chemical Society, 2021. **143**(31): p. 12315–12327.
3. Zhang, D., Atochina-Vasserman, E. N., Maurya, D. S., Liu, M., Xiao, Q., Lu, J., Lauri, G., Ona, N., Reagan, E. K., Ni, H., Weissman, D., & Percec, V., *Targeted Delivery of mRNA with One-Component Ionizable Amphiphilic Janus Dendrimers*. Journal of the American Chemical Society, 2021. **143**(43): p. 17975–17982.
4. Ding, D.Y., Zhang, Y., Jia, Y., Sun, J., *Machine Learning-guided Lipid Nanoparticle Design for mRNA Delivery*. ArXiv, 2023.
5. Maharjan, R., et al., *Recent trends and perspectives of artificial intelligence-based machine learning from discovery to manufacturing in biopharmaceutical industry*. Journal of Pharmaceutical Investigation, 2023. **53**: p. 803–826.
6. Yuan, Y., et al., *Applications of artificial intelligence to lipid nanoparticle delivery*. Particuology, 2024. **90**: p. 88–97.
7. Li, B., et al., *Accelerating ionizable lipid discovery for mRNA delivery using machine learning and combinatorial chemistry*. Nature Materials, 2024. **23**: p. 1002–1008.
8. Xu, Y., et al., *AGILE Platform: A Deep Learning-Powered Approach to Accelerate LNP Development for mRNA Delivery*. bioRxiv, 2023.
9. Lewis, M.M., T.J. Beck, and D. Ghosh, *Applying machine learning to identify ionizable lipids for nanoparticle-mediated delivery of mRNA*. bioRxiv, 2023.
10. Ding, D.Y., et al., *Machine Learning-guided Lipid Nanoparticle Design for mRNA Delivery*. ArXiv, 2023.
11. Kumar, G. and A.M. Ardekani, *Machine-Learning Framework to Predict the Performance of Lipid Nanoparticles for Nucleic Acid Delivery*. ACS applied bio materials, 2024.
12. Bae, S.-H., et al., *Rational Design of Lipid Nanoparticles for Enhanced mRNA Vaccine Delivery via Machine Learning*. Small (Weinheim an Der Bergstrasse, Germany), 2024. **21**.
13. Sato, S., et al., *Understanding the Manufacturing Process of Lipid Nanoparticles for mRNA Delivery Using Machine Learning*. Chemical & pharmaceutical bulletin, 2024. **72** 6: p. 529–539.
14. Hunter, M.R., et al., *Understanding Intracellular Biology to Improve mRNA Delivery by Lipid Nanoparticles*. Small methods, 2023. **7**(9): p. e2201695–n/a.
15. Rafiei, M., A. Shojaei, and Y. Chau, *Machine learning-assisted design of immunomodulatory lipid nanoparticles for delivery of mRNA to repolarize hyperactivated microglia*. Drug Delivery, 2025. **32**.
16. Zhang, D., et al., *Machine learning engineered PoLixNano nanoparticles overcome delivery barriers for nebulized mRNA therapeutics*. bioRxiv, 2024.

17. Kumar, G. and A.M. Ardekani, *Machine-Learning Framework to Predict the Performance of Lipid Nanoparticles for Nucleic Acid Delivery*. ACS Applied Bio Materials, 2025. **8**(5): p. 3717–3727.
18. Zhang, D., Atochina-Vasserman, E. N., Lu, J., Maurya, D. S., Xiao, Q., Liu, M., Adamson, J., Ona, N., Reagan, E. K., Ni, H., Weissman, D., & Percec, V. , *The Unexpected Importance of the Primary Structure of the Hydrophobic Part of One-Component Ionizable Amphiphilic Janus Dendrimers in Targeted mRNA Delivery Activity*. Journal of the American Chemical Society, 2022. **144**(11): p. 4746–4753.
19. Lu, J., Atochina-Vasserman, E. N., Maurya, D. S., Shalihin, M. I., Zhang, D., Chenna, S. S., Adamson, J., Liu, M., Shah, H. U. R., Shah, H., Xiao, Q., Queeley, B., Ona, N. A., Reagan, E. K., Ni, H., Sahoo, D., Peterca, M., Weissman, D., & Percec, V. , *Screening Libraries to Discover Molecular Design Principles for the Targeted Delivery of mRNA with One-Component Ionizable Amphiphilic Janus Dendrimers Derived from Plant Phenolic Acids*. Pharmaceutics, 2023. **15**(6).
20. Sahoo, D., Atochina-Vasserman, E. N., Maurya, D. S., Arshad, M., Chenna, S. S., Ona, N., Vasserman, J. A., Ni, H., Weissman, D., & Percec, V. , *The Constitutional Isomerism of One-Component Ionizable Amphiphilic Janus Dendrimers Orchestrates the Total and Targeted Activities of mRNA Delivery*. Journal of the American Chemical Society, 2024. **146**(6): p. 3627–3634.
21. Sahoo, D., Atochina-Vasserman, E. N., Lu, J., Maurya, D. S., Ona, N., Vasserman, J. A., Ni, H., Berkihsier, S., Park, W.-J., Weissman, D., & Percec, V. , *Toward a Complete Elucidation of the Primary Structure–Activity in Pentaerythritol-Based One-Component Ionizable Amphiphilic Janus Dendrimers for In Vivo Delivery of Luc-mRNA*. Biomacromolecules, 2025. **26**(1): p. 726–737.
22. Chen, T. and C. Guestrin. *XGBoost: A Scalable Tree Boosting System*. in *The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. New York, NY, USA: ACM.
23. NVIDIA. *XGBoost*. 2025 3/14/2025]; Available from: <https://www.nvidia.com/en-us/glossary/xgboost/>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.